|  |  |  |  |
| --- | --- | --- | --- |
| **Project Title:** | **Intelligent Network Intrusion Detection Using Machine Learning** | **Project ID:** |  |
| **Team Members** |  | Date of Request: |  |
|  |  | **Executive Sponsor(s):** | N/A |
|  |  | **Project Type** | N/A |
| **Faculty Sponsor:** |  | **Research Track** |

**Introduction**

Provide a brief overview of your project and explain why your area of research is important. (150 words Max)

|  |
| --- |
| In today’s increasingly digital world, cyber threats pose significant risks to the confidentiality, integrity, and availability of network systems. This project focuses on developing an AI-powered Intrusion Detection System (IDS) that can monitor, analyze, and detect suspicious network activity in real time. Unlike traditional IDS solutions that rely on fixed signatures and predefined rules, this system uses machine learning and deep learning techniques to identify both known and emerging threats. By learning from network behavior and detecting anomalies, the AI-based IDS offers greater accuracy, faster response times, and adaptability to evolving cyberattacks. This project is crucial as it addresses the limitations of conventional security systems and contributes to building smarter, more resilient cybersecurity infrastructures. |

**Problem Definition:**

What problem do you intend to solve in this research? (200 words max)

|  |
| --- |
| Problem Statement:  As cyber threats continue to evolve in complexity and scale, traditional Intrusion Detection Systems (IDSs) that rely on static rules and known attack signatures struggle to detect new, unknown, or subtle attacks. These systems are limited in adaptability and often produce a high rate of false positives, which can overwhelm security teams and lead to delayed or missed threat responses. There is a critical need for an intelligent, adaptive solution that can analyze network traffic in real time and detect both known and emerging threats with high accuracy.  Evidence that the Problem Exists:  Numerous cybersecurity reports show a steady rise in zero-day attacks, advanced persistent threats (APTs), and sophisticated malware that bypass traditional detection systems. Additionally, incidents such as data breaches and ransomware attacks highlight the inability of legacy IDS tools to cope with modern attack vectors. Studies reveal that false positives account for up to 90% of alerts in some systems, reducing efficiency and increasing response times.  Benefits of Solving the Problem:  Developing an AI-based IDS will provide proactive and dynamic threat detection, reduce false alarms, and enhance real-time response. This leads to more robust network security, better resource utilization, and improved protection for sensitive data and infrastructure. |

**Methodology and Experience**

Describe the methodology you intend to use to solve the stated problem. This includes data collection and data analysis. Briefly describe your previous experience with this methodology. If you don't have previous experience, outline your plan to learn this methodology. (200 words max)

|  |
| --- |
| Methodology  The project uses a Machine Learning–based methodology to detect network intrusions. A supervised learning approach is followed, where a classification model is trained on labeled network traffic data. The model learns to distinguish between “normal” and “attack” traffic based on selected network features. After training, the model is integrated with a real-time packet capture system to monitor ongoing traffic and classify packets as safe or malicious.  Data Collection  For training, the **NSL-KDD dataset** is used, which is a well-known benchmark dataset for intrusion detection research. It contains various network connections labeled as either normal or different types of attacks. In the live detection phase, data is collected in real time using **Scapy**, a Python library for packet sniffing and analysis.  Data Analysis  Categorical features such as protocol type, service, and flags are encoded using **Label Encoding**. Numerical features like source bytes and destination bytes are directly included. The processed data is split into training and testing sets. A **Random Forest Classifier** is trained on this data to detect patterns of attacks. Predictions are then validated using metrics like classification reports.  Previous Experience with Methodology or Plans to Learn  I have prior experience working with Python, Scikit-learn, and Pandas for supervised learning tasks such as classification and preprocessing. Although I am still gaining exposure to real-time network packet analysis with Scapy, I am following documentation, tutorials, and small-scale experiments to strengthen my skills in this area. |

**Research Questions and/or Hypotheses:**

Please state you research questions. If you have a hypotheses, also state it here. (150 words max).

|  |
| --- |
| **Research Questions:**   1. Can a machine learning–based model, trained on the NSL-KDD dataset, accurately distinguish between normal and attack network traffic? 2. Which features (such as protocol type, service, flags, and byte counts) contribute most to the detection of intrusions? 3. Can the trained model be effectively integrated with real-time packet capture to detect intrusions as they occur?   **Hypothesis:** A Random Forest classifier trained on labeled network traffic will achieve reliable accuracy in differentiating between normal and malicious packets.  Even with a reduced set of features, the model can still identify attacks with reasonable precision.  Combining offline training with real-time feature extraction will result in a functional intrusion detection system that raises timely alerts for suspicious activity. |

**Literature Review and Related Work (150 words max)**

*Briefly summarize related literature in this area and how your work differs from what has already been done.*

|  |
| --- |
| Intrusion Detection Systems (IDS) have been widely studied using both traditional rule-based techniques and modern machine learning approaches. The KDD’99 and NSL-KDD datasets are among the most commonly used benchmarks for evaluating IDS models. Prior research has explored algorithms such as Support Vector Machines, Decision Trees, and Random Forests to classify normal and attack traffic. More recent studies also focus on deep learning and ensemble methods to improve detection accuracy and reduce false alarms.  While many works remain limited to offline evaluation on benchmark datasets, fewer have combined dataset-driven training with real-time packet analysis. This project differs by not only training a Random Forest classifier on the NSL-KDD dataset but also integrating it with live packet capture using Scapy. This combination bridges the gap between academic IDS research and practical implementation, offering a lightweight system that can raise alerts during active network monitoring. |
|  |

**Deliverables**

What are the deliverables for your project?

|  |
| --- |
| * The key deliverables of this project include: * Trained Machine Learning Model – A Random Forest–based classifier trained on the NSL-KDD dataset, saved for reuse in intrusion detection tasks. * Feature Extraction Module – A Python module capable of extracting essential features (protocol type, service, flags, source bytes, destination bytes) from live network packets. * Real-Time Intrusion Detection System – An integrated system that captures network traffic, applies the trained model, and classifies packets as either normal or attack in real time. * Logging and Alerting Mechanism – A logging system that records intrusion attempts with timestamps and feature details, along with console alerts. * Project Documentation – A comprehensive report detailing the methodology, system design, results, and future improvements. * Source Code Repository – Well-structured Python scripts for training, detection, feature extraction, and packet capturing. |

## Project Team/Roles:

*Who needs to be involved as members of the project team; and what will their roles be on the team? (Please indicate their Gannon student ID number.)*

| Who | User ID | Role on Project Team |
| --- | --- | --- |
|  |  | * Responsible for dataset preparation, feature engineering, and training the Random Forest model. * Implements data preprocessing, model evaluation, and optimization. * Ensures the trained model is integrated with the intrusion detection pipeline. |
|  |  | * Focuses on real-time packet capturing using Scapy and implementing feature extraction from live traffic. * Integrates the detection model with the packet capture system. * Develops the alerting and logging mechanism, and ensures the system runs effectively in real-time conditions. |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |